

# Risk sensitive nonlinear optimal control with measurement uncertainty

Brahayam Ponton\*, Stefan Schaal\*<sup>†</sup> and Ludovic Righetti\*<sup>†</sup>

\*Max-Planck Institute for Intelligent Systems

<sup>†</sup>University of Southern California

**Abstract**—We present an algorithm to synthesize locally-optimal feedback controllers that take into account additive process and measurement noise for nonlinear stochastic optimal control problems. The algorithm is based on an exponential performance criteria which allows to optimize not only the expected value of the cost, but also a linear combination of its higher order moments; thereby, the cost of uncertainty can be taken into account for synthesis of robust or risk-sensitive policies. The method constructs an affine feedback control law, whose gains explicitly depend upon the covariance of the estimation errors and process noise. Despite the fact that controller and observer are designed separately, the measurement noise variance enters the optimal control, therefore generating feedback laws that do not rely on the Certainty Equivalence Principle. The capabilities of the approach are illustrated in simulation on a two degree of freedom (DOF) manipulator, first in a waypoint task and then in a task where the manipulator goes in contact with its environment.

## I. INTRODUCTION

In a not too distant future, personal robots will be a common part of our daily lives, with a broad range of applications going from industrial and service applications to common household scenarios [6]. This can only happen, if they are able to safely operate among humans by being able to *optimally adapt to uncertainty* in a dynamic environment. This is a key ingredient for the future of autonomous robots. In this contribution, we address this aspect by studying the effect of measurement uncertainty in stochastic optimal control problems.

Optimal control theory provides the appropriate framework for tackling the challenge, and in particular, techniques based on *applying Bellman's Principle of Optimality around nominal trajectories* are of significant importance. They blend advantages of local and global methods. On the one hand, they maintain only a single nominal trajectory as a local method; but on the other, they improve it iteratively based on dynamic programming along a tube or neighborhood of the nominal trajectory, what allows them to overcome to some extent the curse of dimensionality and result in computationally efficient algorithms. They approximate the solution of the nonlinear optimal control problem by iteratively solving a first or second order Taylor approximation of the nonlinear problem. Relevant algorithms based on this principle can be found in [5], [7].

The simplifications introduced by the Taylor approximation, that allow to find solutions to the nonlinear stochastic optimal control problem, also impose limitations, namely, they consider only the first moment of the objective function (expected value of a quadratic form), and systems under purely additive noise (Brownian motions with time-dependent diffusion coefficients). As a result, the optimal control law for the stochastic

and deterministic (ignoring noise statistics) optimal control problems are the same. While it is reasonable for systems with small disturbances; intuitively, one would not expect the same to be true for systems with large noise intensity, where a *control strategy capable of reasoning about noise statistics and cost of uncertainty* would be more appropriate.

There exist different ways to address this issue, for example by invalidating the assumptions of the Certainty Equivalence Principle. Ways to achieve this are for instance by considering a nonlinear state equation (as the second derivative of the value function with respect to the state, that enters the diffusion cost of the Hamilton-Jacobi-Bellman (HJB) equation, would no longer be constant), multiplicative noise in the system parameters or by using a non-quadratic objective function (same reason as for a nonlinear state eq.). Care should be taken at preserving the computational efficiency. One of the alternatives, the problem of considering multiplicative process noise has been studied in [13], and extended to multiplicative measurement noise in [4] and [12]. These methods construct an affine control law dependent on noise statistics and successfully apply it to control a two-DOF model of a biomechanical human arm. Another appealing alternative is to be able to capture not only noise effects on the mean of the performance index, but also able to *consider higher order statistics of the performance criteria by using non-quadratic costs*.

Jacobson introduced [2] a risk-sensitive Linear-Exponential-Gaussian (LEG) algorithm, an approach to include the higher order statistics *by using as cost the expectation of the exponential transformation of a performance index*. He showed for a linear system with additive process noise, that the synthesized feedback control law depends explicitly on the noise statistics. For small noise intensity, the LEG control law is similar to the Linear-Quadratic-Gaussian (LQG) controller, but for large noise intensity, they greatly differ. Farshidian and Buchli [1] extended this work for continuous-time stochastic nonlinear optimal control problems. They derived an iterative algorithm, iLEG, for local optimal control and illustrated the effect of considering higher order statistics of the cost in a continuous-cliff problem, where both risk seeking and risk averse control laws can be synthesized.

The extension of the above results to the case where measurement noise is present in the system, is not straightforward. It has been shown [8], that for partially observable systems, because of the multiplicative nature of the exponential cost function, the control law is not anymore a linear functional of the current state, but a functional of the whole smoothed history of states. The proposed solution is the use of an enlarged

state, that grows every timestep to comprise the entire history of states seen so far (use of full-information [9]). Because of this increasingly growing computational complexity, only two special cases (where simplifications in the control law occur) are of interest: when the objective function is a functional only of the final state, and when there is no process noise.

In this paper, we approach the problem from a different perspective in order to reduce these limitations. We will use the idea of sequentially approximating the nonlinear problem by a Taylor expansion and designing risk-sensitive control laws using the exponential transformation of the performance criteria, as recently done in [1]. However, instead of using an enlarged state vector (growing vector composed of the entire history of states), we *use an enlarged dynamical system composed of the dynamics of the control and estimation problems* [14], [3], where the number of states only doubles. In this way, the optimal control law (functional of the state estimate) will be sensitive to both process and measurement noise. From an information perspective, this means that we restrict the amount of information for constructing the optimal control to only statistics that can be captured in the state estimate (we no longer use full-information (entire state sequence)). By doing this, we gain increased flexibility at designing the objective function and are able to capture simultaneously process noise and measurement uncertainty effects on the cost.

The remaining question is *Why does it matter to consider measurement noise effects in the cost and control?* First of all, properly addressing robustness issues due to process, measurement and model parameters noise is an important issue in robotics [6]. More concretely, one can imagine to be performing a reaching motion: in the presence of additive process noise, the motion requires the use of feedback gains proportional to the noise to maintain the performance measured by a cost function (the higher the noise, the higher the gains). In addition, it is natural to trade-off control-effort and cost by allowing some variance in the cost (as it is addressed in [1]). Now, consider the same experiment under measurement noise: you would like to grasp an object using no vision information, or you try to reach a wall to orient yourself in a dark corridor. In these conditions, one behaves more compliant to carefully reach the object (the higher the noise, the lower the gains). This suggests that dynamic interactions of a robot with its environment (a fundamental problem in robotics) could be stated as a problem with measurement uncertainty (location of the contact). Of course, the ability to trade-off performance and control-effort remains important also in this case.

In the following, we first present background material and problem statement. Then we derive the continuous and discrete time algorithms and discuss implementation details. Finally, we illustrate the performance of the controllers in two simple robotics tasks: a way-point and a contact interaction task.

## II. BACKGROUND

In this section, we recall two important results elaborated in [1], [2], [8], and useful for this work: the meaning of the cost's exponential transformation and the form of the HJB eq. under the exponential transformation. Consider the following stochastic nonlinear optimal control problem where the system

dynamics are defined by the stochastic differential eq. (SDE)

$$d\mathbf{x} = \tilde{\mathbf{f}}(\mathbf{x}, \mathbf{u})dt + \tilde{\mathbf{g}}(\mathbf{x}, \mathbf{u})d\tilde{\omega} \quad (1)$$

Let  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{u} \in \mathbb{R}^m$  denote the system states and inputs,  $\tilde{\mathbf{f}}(\mathbf{x}, \mathbf{u})$  and  $\tilde{\mathbf{g}}(\mathbf{x}, \mathbf{u})$  are nonlinear functions representing drift and diffusion coefficients, the Brownian motion  $d\tilde{\omega}$  has zero-mean and covariance  $\Sigma dt$ . The objective function is an exponential transformation of the performance criteria  $\mathcal{J}(\pi)$

$$\mathbf{J} = \min_{\pi} \mathbb{E}\{\exp[\sigma \mathcal{J}(\pi)]\} \quad (2)$$

$\mathcal{J}(\pi)$  is a random variable functional of the policy  $\mathbf{u} = \pi(\mathbf{x})$ .  $\sigma \in \mathbb{R}$  is the risk-sensitive parameter,  $\mathbb{E}$  is the expectation over  $\mathcal{J}(\pi)$ .  $\mathbf{J}$  is therefore the risk-sensitive cost and corresponds to the moment generating function, an alternative specification of the probability distribution of the random variable.

### A. Meaning of the Exponential Transformation of the Cost

It has been shown [1] that the cumulant generating function (logarithmic transformation of the moment generating function) of the risk-sensitive cost can be rewritten as a linear combination of the moments of the performance criteria

$$\frac{1}{\sigma} \log [\mathbf{J}] = \mathbb{E}[\mathcal{J}] + \frac{\sigma}{2} \mu_2[\mathcal{J}] + \frac{\sigma^2}{6} \mu_3[\mathcal{J}] + \dots \quad (3)$$

$\mu_2$ ,  $\mu_3$  denote the second (variance) and third (skewness) moments of  $\mathcal{J}$ . The linear combination depends on the value of the risk-sensitive parameter  $\sigma$ . When  $\sigma$  is zero, the objective function reduces to the mean of the performance criteria  $\mathcal{J}$ . For positive values, all higher order moments have positive coefficients and play the role of a penalty in the cost. This means for example, that there will be a compromise between increasing control effort and narrowing confidence intervals. For negatives values of  $\sigma$ , even moments have negative coefficients and act as a reward. In this case, control effort is reduced by increasing reward (even moments). This can be thought of as playing with the exploration-exploitation trade-off. Higher exploration is achieved, by having lower control-effort. Another interpretation is that strong control effort is avoided in the presence of poor information (high variance).

### B. HJB equation under Exponential Transformation

The second important result, that we would like to recall from [1] is with regard to the form of the HJB eq. under the exponential transformation. Let the performance criteria  $\mathcal{J}$  be

$$\mathcal{J} = \Phi_f(\mathbf{x}_{t_f}) + \int_0^{t_f} L(\mathbf{x}_t, \mathbf{u}_t, t)dt \quad (4)$$

where

$$L(\mathbf{x}_t, \mathbf{u}_t, t) = \Phi(\mathbf{x}_t, t) + \frac{1}{2} \mathbf{u}_t^T \mathbf{R}(\mathbf{x}_t, t) \mathbf{u}_t + \mathbf{u}_t^T \mathbf{r}(\mathbf{x}_t, t) \quad (5)$$

is the rate at which cost increases, composed of a state cost  $\Phi(\mathbf{x}_t, t)$ , which can be nonlinear, and a quadratic control cost.  $\Phi_f(\mathbf{x}_{t_f})$  is the cost at final time.

Under the given dynamics Eq. (1) and cost Eq. (2), the HJB eq. under the exponential transformation takes the form [1]

$$-\partial_t \Psi = \min_{\mathbf{u}_t} \left\{ L + \nabla_{\mathbf{x}} \Psi^T \tilde{\mathbf{f}} + \frac{1}{2} \text{Tr} (\nabla_{\mathbf{x}\mathbf{x}} \Psi \tilde{\mathbf{g}} \Sigma \tilde{\mathbf{g}}^T) + \frac{1}{2} \text{Tr} (\sigma \nabla_{\mathbf{x}} \Psi \nabla_{\mathbf{x}} \Psi^T \tilde{\mathbf{g}} \Sigma \tilde{\mathbf{g}}^T) \right\} \quad (6)$$

where the value function  $\Psi$  is a function of both  $\mathbf{x}$  and  $t$ . The first line is the usual HJB eq. for a stochastic dynamical system with cost rate  $L$  due to the current state and control, the free drift and control benefit costs, and the diffusion cost. The interesting part is the second line term which captures noise effects on statistical properties of the cost (higher moments). When  $\sigma$  is zero, the problem reduces to the minimization of the expected value of the performance criteria  $\mathbb{E}[\mathcal{J}]$ .

### III. PROBLEM FORMULATION

This section introduces the stochastic optimal control problem with measurement noise. Consider the nonlinear dynamical system and measurement model described by the SDEs

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, \mathbf{u})dt + \tilde{\mathbf{F}}(\mathbf{x}, \mathbf{u})d\omega \quad (7)$$

$$d\mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u})dt + \tilde{\mathbf{H}}(\mathbf{x}, \mathbf{u})d\gamma \quad (8)$$

where  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{u} \in \mathbb{R}^m$  and  $\mathbf{y} \in \mathbb{R}^p$  are the system states, control and measured outputs respectively. The Brownian motions  $d\omega$  and  $d\gamma$  are zero-mean with covariance  $\Omega dt$ ,  $\Gamma dt$  respectively.  $\mathbf{f}(\mathbf{x}, \mathbf{u})$  and  $\mathbf{h}(\mathbf{x}, \mathbf{u})$  are the drift coefficients representing the deterministic components of the dynamics and measurement model.  $\tilde{\mathbf{F}}(\mathbf{x}, \mathbf{u})$  and  $\tilde{\mathbf{H}}(\mathbf{x}, \mathbf{u})$  are the diffusion coefficients that encode the stochasticity of the problem.

The cost is given in Eqs. (2), (4) and (5). Our goal is to find a risk-sensitive optimal feedback control law  $\pi^*$  that minimizes the cost for this stochastic system  $\mathbf{J}^\pi(\mathbf{x}_0, t_0)$ , in the presence of additive process noise and measurement uncertainty. Note that the globally optimal control law  $\pi^*(\mathbf{x}, t)$  does not depend on an initial state. However, finding it is in general intractable. Instead, we are interested in a locally-optimal feedback control law that approximates the globally optimal solution in the vicinity of a nominal trajectory  $\mathbf{x}^n(t)$ . Since this nominal trajectory depends on the initial state of the system, so does the optimal feedback control law.

### IV. ALGORITHM DERIVATION

We now derive the continuous and discrete time formulations of the risk-sensitive algorithm. The main idea is to extend the dynamics (7), with the dynamics of an estimator of the state. The optimal feedback control law  $\pi$  becomes a functional of the state estimate and can be iteratively improved, by means of forward and backward passes.

#### A. Continuous time

At each iteration, the algorithm begins with a nominal control sequence  $\mathbf{u}^n(t)$  and the corresponding zero-noise trajectory  $\mathbf{x}^n(t)$ , obtained by applying the control sequence to the dynamics  $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$  with initial state  $\mathbf{x}(0) = \mathbf{x}_0$ .

Next, we form a linear approximation of the dynamics and a quadratic approximation of the cost along the nominal trajectories  $\mathbf{x}^n(t)$  and  $\mathbf{u}^n(t)$ , in terms of state and control

deviations  $\delta\mathbf{x}_t = \mathbf{x}_t - \mathbf{x}^n(t)$ ,  $\delta\mathbf{u}_t = \mathbf{u}_t - \mathbf{u}^n(t)$ . The dynamics and measurement model can then be rewritten as

$$d(\delta\mathbf{x}_t) = (\mathbf{A}_t \delta\mathbf{x}_t + \mathbf{B}_t \delta\mathbf{u}_t)dt + \mathbf{C}_t d\omega_t \quad (9)$$

$$d(\delta\mathbf{y}_t) = (\mathbf{F}_t \delta\mathbf{x}_t + \mathbf{E}_t \delta\mathbf{u}_t)dt + \mathbf{D}_t d\gamma_t \quad (10)$$

where the matrices  $\mathbf{A}_t$ ,  $\mathbf{B}_t$ ,  $\mathbf{C}_t$ ,  $\mathbf{D}_t$ ,  $\mathbf{E}_t$ ,  $\mathbf{F}_t$  are given by (evaluated along the nominal trajectories  $\mathbf{x}^n(t)$ ,  $\mathbf{u}^n(t)$ )

$$\mathbf{A}_t = \partial\mathbf{f}(\mathbf{x}, \mathbf{u})/\partial\mathbf{x}^T, \mathbf{B}_t = \partial\mathbf{f}(\mathbf{x}, \mathbf{u})/\partial\mathbf{u}^T, \mathbf{C}_t = \tilde{\mathbf{F}}(\mathbf{x}, \mathbf{u})$$

$$\mathbf{F}_t = \partial\mathbf{h}(\mathbf{x}, \mathbf{u})/\partial\mathbf{x}^T, \mathbf{E}_t = \partial\mathbf{h}(\mathbf{x}, \mathbf{u})/\partial\mathbf{u}^T, \mathbf{D}_t = \tilde{\mathbf{H}}(\mathbf{x}, \mathbf{u})$$

In the same way, the quadratic approximation of the performance index  $\mathcal{J}$  along the nominal trajectories is given by:

$$\tilde{\ell}(\mathbf{x}, \mathbf{u}, t) = q_t + \mathbf{q}_t^T \delta\mathbf{x}_t + \mathbf{r}_t^T \delta\mathbf{u}_t + \frac{1}{2} \delta\mathbf{x}_t^T \mathbf{Q}_t \delta\mathbf{x}_t + \delta\mathbf{x}_t^T \mathbf{P}_t \delta\mathbf{u}_t + \frac{1}{2} \delta\mathbf{u}_t^T \mathbf{R}_t \delta\mathbf{u}_t \quad (11)$$

$$\tilde{\ell}_f(\mathbf{x}) = q_f + \mathbf{q}_f^T \delta\mathbf{x}_t + \frac{1}{2} \delta\mathbf{x}_t^T \mathbf{Q}_f \delta\mathbf{x}_t \quad (12)$$

We now extend the system dynamics with the dynamics of an estimator using an Extended Kalman filter (EKF)

$$d(\delta\hat{\mathbf{x}}_t) = (\mathbf{A}_t \delta\hat{\mathbf{x}}_t + \mathbf{B}_t \delta\mathbf{u}_t)dt + \mathbf{K}_t [d(\delta\mathbf{y}_t) - d(\delta\hat{\mathbf{y}}_t)] \quad (13)$$

More compactly, the control-estimation dynamics of the enlarged system can be written as:

$$\underbrace{\begin{bmatrix} d(\delta\mathbf{x}_t) \\ d(\delta\hat{\mathbf{x}}_t) \end{bmatrix}}_{d(\delta\tilde{\mathbf{x}}_t)} = \underbrace{\begin{bmatrix} \mathbf{A}_t \delta\mathbf{x}_t + \mathbf{B}_t \delta\mathbf{u}_t \\ \mathbf{A}_t \delta\hat{\mathbf{x}}_t + \mathbf{B}_t \delta\mathbf{u}_t + \mathbf{K}_t \mathbf{F}_t (\delta\mathbf{x}_t - \delta\hat{\mathbf{x}}_t) \end{bmatrix}}_{\tilde{\mathbf{f}}(\delta\tilde{\mathbf{x}}_t, \delta\mathbf{u}_t)} dt + \underbrace{\begin{bmatrix} \mathbf{C}_t & 0 \\ 0 & \mathbf{K}_t \mathbf{D}_t \end{bmatrix}}_{\tilde{\mathbf{g}}(t)} \underbrace{\begin{bmatrix} d\omega_t \\ d\gamma_t \end{bmatrix}}_{\substack{d\omega_t \\ d\gamma_t}} \quad (14)$$

where  $\delta\hat{\mathbf{x}}_t$  is the estimate of  $\delta\mathbf{x}_t$ , and  $\delta\tilde{\mathbf{x}}_t$  represents the vector  $[\delta\mathbf{x}_t, \delta\hat{\mathbf{x}}_t]^T$ . Eq. (14) is almost a linear time-varying (LTV) system (in  $\delta\tilde{\mathbf{x}}_t$ ,  $\delta\mathbf{u}_t$ ,  $d\omega_t$  and  $d\gamma_t$ ), except for the estimation gains  $\mathbf{K}_t$ . If we could fix them, the problem would reduce to a standard LTV. The Principle of Separation of Estimation and Control allows us to do exactly this. We can use a forward pass to pre-compute optimal estimation gains  $\mathbf{K}_t$  along the nominal trajectory with an EKF, and then a backward pass is used to compute the control law. This eases the design of a locally optimal estimator and controller, while still being able to consider the effects of process and measurement uncertainty.

1) *Estimator design:* We present for completeness the procedure for computing the optimal estimation gains  $\mathbf{K}_t$  using a standard EKF in continuous time. Note that other estimators could be used as long as we can extract a sequence of estimation gains. The main idea is to find the gain  $\mathbf{K}_t$  that minimizes the expected outer product of the error dynamics between state and state-estimate dynamics. Error dynamics are

$$\begin{aligned} d(\delta\mathbf{e}_t) &= d(\delta\mathbf{x}_t) - d(\delta\hat{\mathbf{x}}_t) \\ &= \mathbf{A}_t \delta\mathbf{e}_t dt + \mathbf{C}_t d\omega_t - \mathbf{K}_t [d(\delta\mathbf{y}_t) - d(\delta\hat{\mathbf{y}}_t)] \\ &= (\mathbf{A}_t - \mathbf{K}_t \mathbf{F}_t) \delta\mathbf{e}_t dt + \mathbf{C}_t d\omega_t - \mathbf{K}_t \mathbf{D}_t d\gamma_t \end{aligned} \quad (15)$$

To compute the optimal gains, we minimize Eq. (16) with respect to the gains, by taking the first differential with respect

to  $\mathbf{K}_t$  and setting it to zero (Eq. (17)). This results in an optimal estimation gain  $\mathbf{K}_t$  given by Eq. (18).

$$\begin{aligned}\dot{\Sigma}_t^e &= \mathbb{E}[d(\delta \mathbf{e}_t)d(\delta \mathbf{e}_t)^T] \\ &= (\mathbf{A}_t - \mathbf{K}_t \mathbf{F}_t) \Sigma_t^e + \Sigma_t^e (\mathbf{A}_t - \mathbf{K}_t \mathbf{F}_t)^T \\ &\quad + \mathbf{K}_t \mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T + \mathbf{C}_t \Omega_t \mathbf{C}_t^T\end{aligned}\quad (16)$$

$$\begin{aligned}d\dot{\Sigma}_t^e &= d\mathbf{K}_t (\mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T - \mathbf{F}_t \Sigma_t^e) + \\ &\quad (\mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T - \mathbf{F}_t \Sigma_t^e)^T d\mathbf{K}_t^T\end{aligned}\quad (17)$$

$$\mathbf{K}_t = \Sigma_t^e \mathbf{F}_t^T (\mathbf{D}_t \Gamma_t \mathbf{D}_t^T)^{-1} \quad (18)$$

The locally-optimal estimation gains  $\mathbf{K}_t$  computed through a forward pass are then fixed during the backward pass. In this way, the estimation-control system Eq. (14), is in the form of Eq. (1) and is linear in  $\delta \tilde{\mathbf{x}}_t$  and  $\delta \mathbf{u}_t$ , which allows us to make use of the HJB Eq. (6) to compute the control law  $\pi$ .

2) *Controller design:* In this subsection, we show how to compute the optimal feedback control law using a backward pass. The locally-optimal control law will be affine, of the form  $\delta \mathbf{u}_t = \mathbf{l}_t + \mathbf{L}_t \delta \tilde{\mathbf{x}}_t$ . The HJB eq. of the system is given by Eq. (6), using the cost Eqs. (11)-(12) (remember that we use the HJB eq. under the exponential transformation; therefore, the cost need not to be exponentiated), the dynamics Eq. (14) and the quadratic Ansatz for the value function  $\Psi(\delta \tilde{\mathbf{x}}_t, t)$

$$\Psi(\delta \tilde{\mathbf{x}}_t, t) = \frac{1}{2} \begin{bmatrix} \delta \mathbf{x}_t \\ \delta \tilde{\mathbf{x}}_t \end{bmatrix}^T \begin{bmatrix} \mathbf{S}_t^x & \mathbf{S}_t^{x\hat{x}} \\ \mathbf{S}_t^{\hat{x}x} & \mathbf{S}_t^{\hat{x}\hat{x}} \end{bmatrix} \begin{bmatrix} \delta \mathbf{x}_t \\ \delta \tilde{\mathbf{x}}_t \end{bmatrix} + \begin{bmatrix} \delta \mathbf{x}_t \\ \delta \tilde{\mathbf{x}}_t \end{bmatrix}^T \begin{bmatrix} \mathbf{s}_t^x \\ \mathbf{s}_t^{\hat{x}} \end{bmatrix} + s_t$$

where the partial derivatives of the Ansatz  $\Psi$  are given by

$$\begin{aligned}\partial_t \Psi &= \frac{1}{2} \begin{bmatrix} \delta \mathbf{x}_t \\ \delta \tilde{\mathbf{x}}_t \end{bmatrix}^T \begin{bmatrix} \dot{\mathbf{S}}_t^x & \dot{\mathbf{S}}_t^{x\hat{x}} \\ \dot{\mathbf{S}}_t^{\hat{x}x} & \dot{\mathbf{S}}_t^{\hat{x}\hat{x}} \end{bmatrix} \begin{bmatrix} \delta \mathbf{x}_t \\ \delta \tilde{\mathbf{x}}_t \end{bmatrix} + \begin{bmatrix} \delta \mathbf{x}_t \\ \delta \tilde{\mathbf{x}}_t \end{bmatrix}^T \begin{bmatrix} \dot{\mathbf{s}}_t^x \\ \dot{\mathbf{s}}_t^{\hat{x}} \end{bmatrix} + \dot{s}_t \\ \nabla_{\delta \tilde{\mathbf{x}}} \Psi &= \begin{bmatrix} \mathbf{S}_t^x & \mathbf{S}_t^{x\hat{x}} \\ \mathbf{S}_t^{\hat{x}x} & \mathbf{S}_t^{\hat{x}\hat{x}} \end{bmatrix} \begin{bmatrix} \delta \mathbf{x}_t \\ \delta \tilde{\mathbf{x}}_t \end{bmatrix} + \begin{bmatrix} \mathbf{s}_t^x \\ \mathbf{s}_t^{\hat{x}} \end{bmatrix} \\ \nabla_{\delta \tilde{\mathbf{x}} \delta \tilde{\mathbf{x}}} \Psi &= \begin{bmatrix} \mathbf{S}_t^x & \mathbf{S}_t^{x\hat{x}} \\ \mathbf{S}_t^{\hat{x}x} & \mathbf{S}_t^{\hat{x}\hat{x}} \end{bmatrix}\end{aligned}$$

Under the assumed linear dynamics and quadratic cost and value function, we can rewrite the HJB eq. in the following way. The left-hand side (LHS) corresponds to the time derivative of the value function and is given by

$$\begin{aligned}-\frac{1}{2} \delta \mathbf{x}_t^T \dot{\mathbf{S}}_t^x \delta \mathbf{x}_t - \frac{1}{2} \delta \tilde{\mathbf{x}}_t^T \dot{\mathbf{S}}_t^{\hat{x}} \delta \tilde{\mathbf{x}}_t \\ - \delta \mathbf{x}_t^T \dot{\mathbf{S}}_t^{x\hat{x}} \delta \tilde{\mathbf{x}}_t - \delta \tilde{\mathbf{x}}_t^T \dot{\mathbf{S}}_t^{\hat{x}x} \delta \mathbf{x}_t - \dot{s}_t\end{aligned}\quad (19)$$

and the right-hand side (RHS) is the following minimization

$$\begin{aligned} &= \min_{\delta \mathbf{u}_t} \left\{ q_t + \mathbf{q}_t^T \delta \mathbf{x}_t + \mathbf{r}_t^T \delta \mathbf{u}_t + \frac{1}{2} \delta \mathbf{x}_t^T \mathbf{Q}_t \delta \mathbf{x}_t + \delta \mathbf{x}_t^T \mathbf{P}_t \delta \mathbf{u}_t \right. \\ &\quad + \frac{1}{2} \delta \mathbf{u}_t^T \mathbf{R}_t \delta \mathbf{u}_t + (\mathbf{S}_t^x \delta \mathbf{x}_t + \mathbf{S}_t^{x\hat{x}} \delta \tilde{\mathbf{x}}_t + \mathbf{s}_t^x)^T (\mathbf{A}_t \delta \mathbf{x}_t \\ &\quad + \mathbf{B}_t \delta \mathbf{u}_t) + (\mathbf{S}_t^{\hat{x}x} \delta \mathbf{x}_t + \mathbf{S}_t^{\hat{x}\hat{x}} \delta \tilde{\mathbf{x}}_t + \mathbf{s}_t^{\hat{x}})^T (\mathbf{A}_t \delta \tilde{\mathbf{x}}_t \\ &\quad + \mathbf{B}_t \delta \mathbf{u}_t + \mathbf{K}_t \mathbf{F}_t (\delta \mathbf{x}_t - \delta \tilde{\mathbf{x}}_t)) + \frac{\sigma}{2} (\mathbf{S}_t^x \delta \mathbf{x}_t \\ &\quad + \mathbf{S}_t^{x\hat{x}} \delta \tilde{\mathbf{x}}_t + \mathbf{s}_t^x)^T \mathbf{C}_t \Omega_t \mathbf{C}_t^T (\mathbf{S}_t^x \delta \mathbf{x}_t + \mathbf{S}_t^{x\hat{x}} \delta \tilde{\mathbf{x}}_t + \mathbf{s}_t^x) \\ &\quad + \frac{\sigma}{2} (\mathbf{S}_t^{\hat{x}x} \delta \mathbf{x}_t + \mathbf{S}_t^{\hat{x}\hat{x}} \delta \tilde{\mathbf{x}}_t + \mathbf{s}_t^{\hat{x}})^T \mathbf{K}_t \mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T \\ &\quad \times (\mathbf{S}_t^{x\hat{x}} \delta \mathbf{x}_t + \mathbf{S}_t^{\hat{x}\hat{x}} \delta \tilde{\mathbf{x}}_t + \mathbf{s}_t^{\hat{x}}) + \frac{1}{2} \text{Tr} (\mathbf{S}_t^x \mathbf{C}_t \Omega_t \mathbf{C}_t^T) \\ &\quad \left. + \frac{1}{2} \text{Tr} (\mathbf{S}_t^{\hat{x}} \mathbf{K}_t \mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T) \right\} \quad (20)\end{aligned}$$

In order to perform the minimization, we analyze the control dependent terms in the above expression, which are given by

$$\begin{aligned}V_{\delta \mathbf{u}_t} &= \frac{1}{2} \delta \mathbf{u}_t^T \underbrace{\mathbf{R}_t}_{\mathbf{H}_t} \delta \mathbf{u}_t + \delta \mathbf{u}_t^T \underbrace{(\mathbf{r}_t + \mathbf{B}_t^T (\mathbf{s}_t^x + \mathbf{s}_t^{\hat{x}}))}_{\mathbf{g}_t} \\ &\quad + \underbrace{(\mathbf{P}_t^T + \mathbf{B}_t^T (\mathbf{S}_t^x + \mathbf{S}_t^{x\hat{x}}))}_{\mathbf{G}_t^x} \delta \mathbf{x}_t + \underbrace{\mathbf{B}_t^T (\mathbf{S}_t^{x\hat{x}} + \mathbf{S}_t^{\hat{x}})}_{\mathbf{G}_t^{\hat{x}}} \delta \tilde{\mathbf{x}}_t\end{aligned}$$

The above expression is quadratic in  $\delta \mathbf{u}_t$  and is easy to minimize. However, the minimum is a functional not only of  $\delta \tilde{\mathbf{x}}_t$ , but also of  $\delta \mathbf{x}_t$ , as we expected [8],[9]. At this point, we use the assumption that we do not have access to full state information, only a statistical description of it, given by the state estimate. Therefore, in order to perform the minimization, we take an expectation of  $V_{\delta \mathbf{u}_t}$  over  $\delta \mathbf{x}_t$  conditioned on  $\delta \tilde{\mathbf{x}}_t$

$$\mathbb{E}_{\delta \mathbf{x}_t | \delta \tilde{\mathbf{x}}_t} [V_{\delta \mathbf{u}_t}] = \frac{1}{2} \delta \mathbf{u}_t^T \mathbf{H}_t \delta \mathbf{u}_t + \delta \mathbf{u}_t^T (\mathbf{g}_t + (\mathbf{G}_t^x + \mathbf{G}_t^{\hat{x}}) \delta \tilde{\mathbf{x}}_t)$$

This means that the cost of uncertainty due to measurement noise, considers only the effects of mean and variance of the measurement (captured by the EKF) when evaluating noise effects on the statistical properties of the performance criteria. Consequently, the computed risk-sensitive control law, considers only as cost of uncertainty the one that can be computed by means of the state estimate.

From the above expression, the minimizer can be analytically computed or in case of control constraints, a quadratic program can be used to solve for the constrained minimizer [11]. In both cases, the minimizer is an affine functional of the state-estimate. For the unconstrained case, it is given by

$$\delta \mathbf{u}_t = \mathbf{l}_t + \mathbf{L}_t \delta \tilde{\mathbf{x}}_t = -\mathbf{H}_t^{-1} \mathbf{g}_t - \mathbf{H}_t^{-1} (\mathbf{G}_t^x + \mathbf{G}_t^{\hat{x}}) \delta \tilde{\mathbf{x}}_t \quad (21)$$

The control dependent terms  $V_{\delta \mathbf{u}_t}$  can then be written in terms of the optimal control as:

$$\begin{aligned}V_{\delta \mathbf{u}_t^*} &= \frac{1}{2} \delta \tilde{\mathbf{x}}_t^T ((\mathbf{G}_t^x)^T \mathbf{H}_t^{-1} \mathbf{G}_t^x - (\mathbf{G}_t^{\hat{x}})^T \mathbf{H}_t^{-1} \mathbf{G}_t^{\hat{x}}) \delta \tilde{\mathbf{x}}_t - \\ &\quad \delta \mathbf{x}_t^T ((\mathbf{G}_t^x)^T \mathbf{H}_t^{-1} (\mathbf{G}_t^x + \mathbf{G}_t^{\hat{x}})) \delta \tilde{\mathbf{x}}_t - \frac{1}{2} \mathbf{g}_t^T \mathbf{H}_t^{-1} \mathbf{g}_t - \\ &\quad \delta \mathbf{x}_t^T (\mathbf{G}_t^x)^T \mathbf{H}_t^{-1} \mathbf{g}_t - \delta \tilde{\mathbf{x}}_t^T (\mathbf{G}_t^{\hat{x}})^T \mathbf{H}_t^{-1} \mathbf{g}_t\end{aligned}\quad (22)$$

The negative coefficients in the terms of  $V_{\delta \mathbf{u}_t^*}$  are the benefit of control at reducing the cost. It should be noted, that even setting measurement noise to zero does not give a control law equivalent to what was found in [1]. It should be clear from Eq. (21) that mathematically they are not the same. But intuitively, this control law minimizes the cost of uncertainty due to process and measurement noise. Given that there is no measurement noise, there is more control authority for reducing cost due to process noise. Writing these terms back into the RHS of the HJB, we can drop the minimization and verify that the quadratic Ansatz for the value function remains quadratic and is therefore valid. This is not the case, if we explicitly considered multiplicative uncertainty. Finally, matching terms in LHS and RHS of the HJB eq., we can write

the backward pass recursion eqs. as

$$\begin{aligned}
-\dot{\mathbf{S}}_t^x &= \mathbf{Q}_t + \mathbf{A}_t^T \mathbf{S}_t^x + (\mathbf{S}_t^x)^T \mathbf{A}_t + \mathbf{S}_t^{x\hat{x}} \mathbf{K}_t \mathbf{F}_t + \mathbf{F}_t^T \mathbf{K}_t^T (\mathbf{S}_t^{x\hat{x}})^T \\
&\quad + \sigma (\mathbf{S}_t^x)^T \mathbf{C}_t \Omega_t \mathbf{C}_t^T \mathbf{S}_t^x + \sigma \mathbf{S}_t^{x\hat{x}} \mathbf{K}_t \mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T (\mathbf{S}_t^{x\hat{x}})^T \\
-\dot{\mathbf{S}}_t^{\hat{x}} &= (\mathbf{A}_t - \mathbf{K}_t \mathbf{F}_t)^T \mathbf{S}_t^{\hat{x}} + (\mathbf{S}_t^{\hat{x}})^T (\mathbf{A}_t - \mathbf{K}_t \mathbf{F}_t) \\
&\quad + (\mathbf{G}_t^x)^T \mathbf{H}_t^{-1} \mathbf{G}_t^x - (\mathbf{G}_t^{\hat{x}})^T \mathbf{H}_t^{-1} \mathbf{G}_t^{\hat{x}} \\
&\quad + \sigma (\mathbf{S}_t^{x\hat{x}})^T \mathbf{C}_t \Omega_t \mathbf{C}_t^T \mathbf{S}_t^{x\hat{x}} + \sigma (\mathbf{S}_t^{\hat{x}})^T \mathbf{K}_t \mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T \mathbf{S}_t^{\hat{x}} \\
-\dot{\mathbf{S}}_t^{x\hat{x}} &= \mathbf{A}_t^T \mathbf{S}_t^{x\hat{x}} + \mathbf{S}_t^{x\hat{x}} (\mathbf{A}_t - \mathbf{K}_t \mathbf{F}_t) + \mathbf{F}_t^T \mathbf{K}_t^T \mathbf{S}_t^{\hat{x}} \\
&\quad - (\mathbf{G}_t^x)^T \mathbf{H}_t^{-1} (\mathbf{G}_t^x + \mathbf{G}_t^{\hat{x}}) \\
&\quad + \sigma (\mathbf{S}_t^x)^T \mathbf{C}_t \Omega_t \mathbf{C}_t^T \mathbf{S}_t^{x\hat{x}} + \sigma \mathbf{S}_t^{x\hat{x}} \mathbf{K}_t \mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T \mathbf{S}_t^x \\
-\dot{\mathbf{s}}_t^x &= \mathbf{q}_t + \mathbf{A}_t^T \mathbf{s}_t^x + \mathbf{F}_t^T \mathbf{K}_t^T \mathbf{s}_t^{\hat{x}} - (\mathbf{G}_t^x)^T \mathbf{H}_t^{-1} \mathbf{g}_t \\
&\quad + \sigma (\mathbf{S}_t^x)^T \mathbf{C}_t \Omega_t \mathbf{C}_t^T \mathbf{s}_t^x + \sigma \mathbf{S}_t^{x\hat{x}} \mathbf{K}_t \mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T \mathbf{s}_t^{\hat{x}} \\
-\dot{\mathbf{s}}_t^{\hat{x}} &= (\mathbf{A}_t - \mathbf{K}_t \mathbf{F}_t)^T \mathbf{s}_t^{\hat{x}} - (\mathbf{G}_t^{\hat{x}})^T \mathbf{H}_t^{-1} \mathbf{g}_t \\
&\quad + \sigma (\mathbf{S}_t^{x\hat{x}})^T \mathbf{C}_t \Omega_t \mathbf{C}_t^T \mathbf{s}_t^x + \sigma (\mathbf{S}_t^{\hat{x}})^T \mathbf{K}_t \mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T \mathbf{s}_t^{\hat{x}} \\
-\dot{\mathbf{s}}_t &= q_t - \frac{1}{2} \mathbf{g}_t^T \mathbf{H}_t^{-1} \mathbf{g}_t + \frac{1}{2} \text{Tr}(\mathbf{S}_t^x \mathbf{C}_t \Omega_t \mathbf{C}_t^T) \\
&\quad + \frac{1}{2} \text{Tr}(\mathbf{S}_t^{\hat{x}} \mathbf{K}_t \mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T) + \frac{\sigma}{2} (\mathbf{s}_t^x)^T \mathbf{C}_t \Omega_t \mathbf{C}_t^T \mathbf{s}_t^x \\
&\quad + \frac{\sigma}{2} (\mathbf{s}_t^{\hat{x}})^T \mathbf{K}_t \mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T \mathbf{s}_t^{\hat{x}} \quad (23)
\end{aligned}$$

The integration runs backward in time with  $\mathbf{S}_t^x = \mathbf{Q}_f$ ,  $\mathbf{S}_t^{\hat{x}} = 0$ ,  $\mathbf{S}_t^{x\hat{x}} = 0$ ,  $\mathbf{s}_t^x = \mathbf{q}_f$ ,  $\mathbf{s}_t^{\hat{x}} = 0$  and  $s_t = q_f$ .

**Remark** The effects of process and measurement noise appear in pairs, this is due to the fact that we assumed their Brownian motions to be uncorrelated (see  $\tilde{\mathbf{g}}(t)$  Eq. (14)). However, their combined effect is not just as having higher process noise. Estimation couples their effects (Eq. (15)), and this can be seen in the recursion eqs., where we do not only have costs for the state and its estimate  $\mathbf{S}_t^x$  and  $\mathbf{S}_t^{\hat{x}}$ , but also the coupling cost  $\mathbf{S}_t^{x\hat{x}}$ ; whose products with the covariances of process noise and estimation error determine how process noise and measurement uncertainty affect the value function and therefore the control law. As will be seen in the experimental section, the higher the process noise, the higher the feedback gains; while the opposite holds for measurement noise.

### B. Discrete time

We will now briefly derive a discrete-time version of the algorithm, more amenable for computational implementation and control, but which remains similar in spirit to the continuous time version. Note that we skip all the details that can easily be recovered from the continuous case discussion.

The algorithm begins with a nominal control  $\mathbf{u}_k^n$  and state sequence  $\mathbf{x}_k^n$ . Next, we discretize and linearize the dynamics and quadratize the cost along  $\mathbf{u}_k^n$ ,  $\mathbf{x}_k^n$  in terms of state and control deviations  $\delta \mathbf{x}_k = \mathbf{x}_k - \mathbf{x}_k^n$ ,  $\delta \mathbf{u}_k = \mathbf{u}_k - \mathbf{u}_k^n$ , obtaining:

$$\delta \mathbf{x}_{k+1} = \mathbf{A}_k \delta \mathbf{x}_k + \mathbf{B}_k \delta \mathbf{u}_k + \mathbf{C}_k \omega_k \quad (24)$$

$$\delta \mathbf{y}_{k+1} = \mathbf{F}_k \delta \mathbf{x}_k + \mathbf{E}_k \delta \mathbf{u}_k + \mathbf{D}_k \gamma_k \quad (25)$$

where  $\omega_k \sim \mathcal{N}(0, \Omega_k)$  and  $\gamma_k \sim \mathcal{N}(0, \Gamma_k)$ . The matrices  $\mathbf{A}_k$ ,  $\mathbf{B}_k$ ,  $\mathbf{C}_k$ ,  $\mathbf{D}_k$ ,  $\mathbf{E}_k$ ,  $\mathbf{F}_k$  are given by:

$$\begin{aligned}
\mathbf{A}_k &= \mathbf{I} + \Delta t \partial \mathbf{f} / \partial \mathbf{x}^T, \quad \mathbf{B}_k = \Delta t \partial \mathbf{f} / \partial \mathbf{u}^T, \quad \mathbf{C}_k = \sqrt{\Delta t} \tilde{\mathbf{F}}(\mathbf{x}, \mathbf{u}) \\
\mathbf{F}_k &= \Delta t \partial \mathbf{h} / \partial \mathbf{x}^T, \quad \mathbf{E}_k = \Delta t \partial \mathbf{h} / \partial \mathbf{u}^T, \quad \mathbf{D}_k = \sqrt{\Delta t} \tilde{\mathbf{H}}(\mathbf{x}, \mathbf{u})
\end{aligned}$$

The quadratic approximation of the performance index  $\mathcal{J}$  is:

$$\begin{aligned}
\tilde{\ell}(\mathbf{x}, \mathbf{u}, k) &= q_k + \mathbf{q}_k^T \delta \mathbf{x}_k + \mathbf{r}_k^T \delta \mathbf{u}_k + \frac{1}{2} \delta \mathbf{x}_k^T \mathbf{Q}_k \delta \mathbf{x}_k + \\
&\quad \delta \mathbf{x}_k^T \mathbf{P}_k \delta \mathbf{u}_k + \frac{1}{2} \delta \mathbf{u}_k^T \mathbf{R}_k \delta \mathbf{u}_k \quad (26)
\end{aligned}$$

$$\tilde{\ell}_N(\mathbf{x}) = q_N + \mathbf{q}_N^T \delta \mathbf{x}_k + \frac{1}{2} \delta \mathbf{x}_k^T \mathbf{Q}_N \delta \mathbf{x}_k \quad (27)$$

where  $q_k = q_t \Delta t$ ,  $\mathbf{q}_k = \mathbf{q}_t \Delta t$ ,  $\mathbf{r}_k = \mathbf{r}_t \Delta t$ ,  $\mathbf{Q}_k = \mathbf{Q}_t \Delta t$ ,  $\mathbf{P}_k = \mathbf{P}_t \Delta t$  and  $\mathbf{R}_k = \mathbf{R}_t \Delta t$ . The dynamics of the composed control and estimation problems, is similar to Eq. (14).

1) *Estimator design:* We use an EKF, that minimizes the expected outer-product of the error dynamics. The optimal estimation gains that minimize

$$\begin{aligned}
\Sigma_{k+1}^e &= (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k) \Sigma_k^e (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k)^T \\
&\quad + \mathbf{K}_k \mathbf{D}_k \Gamma_k \mathbf{D}_k^T \mathbf{K}_k^T + \mathbf{C}_k \Omega_k \mathbf{C}_k^T \quad (28)
\end{aligned}$$

are given by

$$\mathbf{K}_k = \mathbf{A}_k \Sigma_k^e \mathbf{F}_k^T (\mathbf{F}_t \Sigma_k^e \mathbf{F}_t^T + \mathbf{D}_t \Gamma_t \mathbf{D}_t^T)^{-1} \quad (29)$$

They are updated at each iteration in a forward pass along the nominal trajectories, and are then fixed for the backward pass,

2) *Controller design:* The locally-optimal control law is affine, of the form  $\delta \mathbf{u}_k = \mathbf{l}_k + \mathbf{L}_k \delta \hat{\mathbf{x}}_k$ . The Ansatz for the value function is still quadratic and the recursion eq. is

$$\Psi_\sigma(\delta \hat{\mathbf{x}}_k, k) = \min_{\mathbf{u}} \left\{ \tilde{\ell}(\mathbf{x}, \mathbf{u}, k) + \mathbb{E}[\Psi_\sigma(\delta \hat{\mathbf{x}}_{k+1}, k+1)] \right\}$$

The subscript  $\sigma$  is to remind us that, although this eq. is similar to the usual Bellman eq., noise propagation is different. Here, besides the usual diffusion cost (noise effects on the mean), the cost of uncertainty given by the HJB Eq. (6) in the  $\sigma$ -dependent term is included. Using the linear dynamics of the enlarged system, the quadratic cost Eqs. (26)-(27), and the quadratic Ansatz, it is easy, but long to write the Bellman eq., which is why we omit it here. It is worth pointing out, that the gradients of the Ansatz for evaluating how noise propagates, should not be evaluated at  $\delta \hat{\mathbf{x}}_k$ , but at  $\delta \hat{\mathbf{x}}_{k+1}$ . This allows to capture noise effects in the control terms, and derive a risk-sensitive control law. The control law has the form

$$\delta \mathbf{u}_k = \mathbf{l}_k + \mathbf{L}_k \delta \hat{\mathbf{x}}_k = -\mathbf{H}_k^{-1} \mathbf{g}_k - \mathbf{H}_k^{-1} (\mathbf{G}_k^x + \mathbf{G}_k^{\hat{x}}) \delta \hat{\mathbf{x}}_k \quad (30)$$

where  $\mathbf{H}_k$ ,  $\mathbf{g}_k$ ,  $\mathbf{G}_k^x$  and  $\mathbf{G}_k^{\hat{x}}$  are given by

$$\begin{aligned}
\mathbf{H}_k &= \mathbf{R}_k + \mathbf{B}_k^T (\mathbf{S}_k^x + \mathbf{S}_k^{\hat{x}} + \mathbf{S}_k^{x\hat{x}} + \mathbf{S}_k^{\hat{x}x}) \mathbf{B}_k + \\
&\quad \sigma \mathbf{B}_k^T (\mathbf{S}_k^x + \mathbf{S}_k^{x\hat{x}})^T \mathbf{C}_k \Omega_k \mathbf{C}_k^T (\mathbf{S}_k^x + \mathbf{S}_k^{x\hat{x}}) \mathbf{B}_k + \\
&\quad \sigma \mathbf{B}_k^T (\mathbf{S}_k^{\hat{x}x} + \mathbf{S}_k^{\hat{x}})^T \mathbf{K}_k \mathbf{D}_k \Gamma_k \mathbf{D}_k^T \mathbf{K}_k^T (\mathbf{S}_k^{\hat{x}x} + \mathbf{S}_k^{\hat{x}}) \mathbf{B}_k \\
\mathbf{g}_k &= \mathbf{r}_k + \mathbf{B}_k^T (\mathbf{s}_k^x + \mathbf{s}_k^{\hat{x}}) + \sigma \mathbf{B}_k^T (\mathbf{S}_k^x + \mathbf{S}_k^{x\hat{x}})^T \mathbf{C}_k \Omega_k \mathbf{C}_k^T \mathbf{s}_k^x + \\
&\quad \sigma \mathbf{B}_k^T (\mathbf{S}_k^{\hat{x}x} + \mathbf{S}_k^{\hat{x}})^T \mathbf{K}_k \mathbf{D}_k \Gamma_k \mathbf{D}_k^T \mathbf{K}_k^T \mathbf{s}_k^{\hat{x}} \\
\mathbf{G}_k^x &= \mathbf{P}_k^T + \mathbf{B}_k^T (\mathbf{S}_k^x + \mathbf{S}_k^{x\hat{x}}) \mathbf{A}_k + \mathbf{B}_k^T (\mathbf{S}_k^{\hat{x}} + \mathbf{S}_k^{x\hat{x}}) \mathbf{K}_k \mathbf{F}_k + \\
&\quad \sigma \mathbf{B}_k^T (\mathbf{S}_k^x + \mathbf{S}_k^{x\hat{x}})^T \mathbf{C}_k \Omega_k \mathbf{C}_k^T (\mathbf{S}_k^x \mathbf{A}_k + \mathbf{S}_k^{x\hat{x}} \mathbf{K}_k \mathbf{F}_k) + \\
&\quad \sigma \mathbf{B}_k^T (\mathbf{S}_k^{\hat{x}x} + \mathbf{S}_k^{\hat{x}})^T \mathbf{K}_k \mathbf{D}_k \Gamma_k \mathbf{D}_k^T \mathbf{K}_k^T (\mathbf{S}_k^{\hat{x}x} \mathbf{A}_k + \mathbf{S}_k^{\hat{x}} \mathbf{K}_k \mathbf{F}_k) \\
\mathbf{G}_k^{\hat{x}} &= \mathbf{B}_k^T (\mathbf{S}_k^{\hat{x}} + \mathbf{S}_k^{x\hat{x}}) (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k) + \\
&\quad \sigma \mathbf{B}_k^T (\mathbf{S}_k^x + \mathbf{S}_k^{x\hat{x}})^T \mathbf{C}_k \Omega_k \mathbf{C}_k^T \mathbf{S}_k^{x\hat{x}} (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k) + \\
&\quad \sigma \mathbf{B}_k^T (\mathbf{S}_k^{\hat{x}x} + \mathbf{S}_k^{\hat{x}})^T \mathbf{K}_k \mathbf{D}_k \Gamma_k \mathbf{D}_k^T \mathbf{K}_k^T \mathbf{S}_k^{\hat{x}} (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k)
\end{aligned}$$

The control dependent terms, as a function of the minimizer, take the same form as in Eq. (22), but in discrete time. Finally, by reinserting the terms  $V_{\delta \mathbf{u}_k}$  in the RHS of the Bellman eq., and grouping together terms with similar coefficients of  $\delta \mathbf{x}_k$  and  $\delta \hat{\mathbf{x}}_k$ , we get the following set of backward recursion eqs.

$$\begin{aligned}
\mathbf{S}_k^x &= \mathbf{Q}_k + \mathbf{A}_k^T \mathbf{S}_{k+1}^x \mathbf{A}_k + (\mathbf{F}_k^T \mathbf{K}_k^T \mathbf{S}_{k+1}^{\hat{x}} + 2\mathbf{A}_k^T \mathbf{S}_{k+1}^{x\hat{x}}) \mathbf{K}_k \mathbf{F}_k \\
&\quad + \sigma(\mathbf{S}_{k+1}^x \mathbf{A}_k + \mathbf{S}_{k+1}^{x\hat{x}} \mathbf{K}_k \mathbf{F}_k)^T \mathbf{C}_k \Omega_k \mathbf{C}_k^T (\mathbf{S}_{k+1}^x \mathbf{A}_k \\
&\quad + \mathbf{S}_{k+1}^{x\hat{x}} \mathbf{K}_k \mathbf{F}_k) + \sigma((\mathbf{S}_{k+1}^{\hat{x}})^T \mathbf{A}_k + \mathbf{S}_{k+1}^{\hat{x}x} \mathbf{K}_k \mathbf{F}_k)^T \mathbf{K}_k \mathbf{D}_k \Gamma_k \\
&\quad \times \mathbf{D}_k^T \mathbf{K}_k^T ((\mathbf{S}_{k+1}^{\hat{x}})^T \mathbf{A}_k + \mathbf{S}_{k+1}^{\hat{x}x} \mathbf{K}_k \mathbf{F}_k) \\
\mathbf{S}_k^{\hat{x}} &= (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k)^T \mathbf{S}_{k+1}^{\hat{x}} (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k) + (\mathbf{G}_k^x + \mathbf{G}_k^{\hat{x}})^T \\
&\quad \times \mathbf{H}_k^{-1} (\mathbf{G}_k^x - \mathbf{G}_k^{\hat{x}}) + \sigma(\mathbf{S}_{k+1}^{\hat{x}} (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k))^T \mathbf{C}_k \Omega_k \mathbf{C}_k^T \\
&\quad \times \mathbf{S}_{k+1}^{\hat{x}} (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k) + \sigma(\mathbf{S}_{k+1}^{\hat{x}} (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k))^T \mathbf{K}_k \mathbf{D}_k \\
&\quad \times \Gamma_k \mathbf{D}_k^T \mathbf{K}_k^T \mathbf{S}_{k+1}^{\hat{x}} (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k) \\
\mathbf{S}_k^{x\hat{x}} &= ((\mathbf{S}_{k+1}^{\hat{x}})^T \mathbf{A}_k + (\mathbf{S}_{k+1}^{\hat{x}x})^T \mathbf{K}_k \mathbf{F}_k)^T (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k) - (\mathbf{G}_k^x)^T \\
&\quad \times \mathbf{H}_k^{-1} (\mathbf{G}_k^x + \mathbf{G}_k^{\hat{x}}) \\
&\quad + \sigma(\mathbf{S}_{k+1}^x \mathbf{A}_k + \mathbf{S}_{k+1}^{x\hat{x}} \mathbf{K}_k \mathbf{F}_k)^T \mathbf{C}_k \Omega_k \mathbf{C}_k^T \mathbf{S}_{k+1}^{x\hat{x}} (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k) \\
&\quad + \sigma((\mathbf{S}_{k+1}^{\hat{x}})^T \mathbf{A}_k + \mathbf{S}_{k+1}^{\hat{x}x} \mathbf{K}_k \mathbf{F}_k)^T \mathbf{K}_k \mathbf{D}_k \Gamma_k \mathbf{D}_k^T \mathbf{K}_k^T \mathbf{S}_{k+1}^{\hat{x}} \\
&\quad \times (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k) \\
\mathbf{s}_k^x &= \mathbf{q}_k + \mathbf{A}_k^T \mathbf{s}_{k+1}^x + \mathbf{F}_k^T \mathbf{K}_k^T \mathbf{s}_{k+1}^{\hat{x}} - (\mathbf{G}_k^x)^T \mathbf{H}_k^{-1} \mathbf{g}_k \\
&\quad + \sigma(\mathbf{S}_{k+1}^x \mathbf{A}_k + \mathbf{S}_{k+1}^{x\hat{x}} \mathbf{K}_k \mathbf{F}_k)^T \mathbf{C}_k \Omega_k \mathbf{C}_k^T \mathbf{s}_{k+1}^x \\
&\quad + \sigma((\mathbf{S}_{k+1}^{\hat{x}})^T \mathbf{A}_k + \mathbf{S}_{k+1}^{\hat{x}x} \mathbf{K}_k \mathbf{F}_k)^T \mathbf{K}_k \mathbf{D}_k \Gamma_k \mathbf{D}_k^T \mathbf{K}_k^T \mathbf{s}_{k+1}^{\hat{x}} \\
\mathbf{s}_k^{\hat{x}} &= (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k)^T \mathbf{s}_{k+1}^{\hat{x}} - (\mathbf{G}_k^{\hat{x}})^T \mathbf{H}_k^{-1} \mathbf{g}_k \\
&\quad + \sigma(\mathbf{S}_{k+1}^{\hat{x}} (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k))^T \mathbf{C}_k \Omega_k \mathbf{C}_k^T \mathbf{s}_{k+1}^{\hat{x}} \\
&\quad + \sigma(\mathbf{S}_{k+1}^{\hat{x}} (\mathbf{A}_k - \mathbf{K}_k \mathbf{F}_k))^T \mathbf{K}_k \mathbf{D}_k \Gamma_k \mathbf{D}_k^T \mathbf{K}_k^T \mathbf{s}_{k+1}^{\hat{x}} \\
\mathbf{s}_k &= \mathbf{s}_{k+1} + \mathbf{q}_k - \frac{1}{2} \mathbf{g}_k^T \mathbf{H}_k^{-1} \mathbf{g}_k + \frac{1}{2} \text{Tr} \left( \mathbf{S}_{k+1}^x \mathbf{C}_k \Omega_k \mathbf{C}_k^T \right) \\
&\quad + \frac{1}{2} \text{Tr} \left( \mathbf{S}_{k+1}^{\hat{x}} \mathbf{K}_k \mathbf{D}_k \Gamma_k \mathbf{D}_k^T \mathbf{K}_k^T \right) + \frac{\sigma}{2} (\mathbf{s}_{k+1}^x)^T \mathbf{C}_k \Omega_k \mathbf{C}_k^T \mathbf{s}_{k+1}^x \\
&\quad + \frac{\sigma}{2} (\mathbf{s}_{k+1}^{\hat{x}})^T \mathbf{K}_k \mathbf{D}_k \Gamma_k \mathbf{D}_k^T \mathbf{K}_k^T \mathbf{s}_{k+1}^{\hat{x}} \quad (31)
\end{aligned}$$

where the recursion runs backward in time from  $\mathbf{S}_N^x = \mathbf{Q}_N$ ,  $\mathbf{S}_N^{\hat{x}} = 0$ ,  $\mathbf{S}_N^{x\hat{x}} = 0$ ,  $\mathbf{s}_N^x = \mathbf{q}_N$ ,  $\mathbf{s}_N^{\hat{x}} = 0$  and  $s_N = q_N$ .

### C. Implementation details and algorithm summary

#### Algorithm 1 Risk-Sensitive Nonlinear Control

##### Given:

- Cost function (4), system dynamics (7) and measurement model (8)
- Risk sensitivity parameter  $\sigma$

##### Initialization:

- Start with a stable control law  $\pi(t, \mathbf{x})$

##### repeat

- Forward integrate or propagate the system dynamics to compute a nominal trajectory:  $\mathbf{x}_{1:T}^n, \mathbf{u}_{1:T}^n$
- Linear approximation of the dynamics. Eqs. (9) or (24)
- Quadratic approximation of the cost. Eqs. (10) or (25)
- Forward pass for estimation gains. Eqs. (16), (18) or (28), (29)
- Backward pass with regularization parameter  $\lambda$ . Eqs. (23) or (31)
- Update control law with line search parameter  $\alpha$ ,  $\pi(t, \mathbf{x}) = \mathbf{u}^n(t) + \alpha \mathbf{l}(t) + \mathbf{L}(t)(\mathbf{x}(t) - \mathbf{x}^n(t))$

until convergence or a termination condition is satisfied

Algorithm 1 summarizes the procedure to compute locally-optimal feedback control policies sensitive to process and measurement noise. Here, we mention two implementation details for the discrete-time case, that we use, as presented in [10]. First of all, when computing the optimal control

in Eq. (30),  $\mathbf{H}_k$  needs to be inverted. When it is positive-definite, the unique minimizer can be readily found. If it is not, there exist control directions which would allow to make the cost arbitrarily small. This is obviously not true for the nonlinear system, it appears because of the approximations (compare for example  $\mathbf{H}_k$  with  $\mathbf{H}_t$ ). To control this, we use a regularization term  $\lambda \mathbf{I}$  that makes the sum  $\mathbf{H}_k + \lambda \mathbf{I}$  positive-definite. In our case, this brings an additional advantage, when the sensitivity parameter takes a huge negative value, this can also make the matrix  $\mathbf{H}_k$  be negative-definite and the feedback gains too loose. By including the regularization term, this does not happen anymore. This is in contrast with [1], where  $\mathbf{H}_t$  cannot capture noise effects. An outer-loop regulates  $\lambda$ , similar to [10]. The second detail, is that when propagating the dynamics with the updated control sequence, the new state trajectory might diverge. By adding a line search with parameter  $\alpha \in [0, 1]$ , a control sequence that generates a reduction in the cost can still be found and progress in the optimization can be made.

### V. EXPERIMENTAL RESULTS

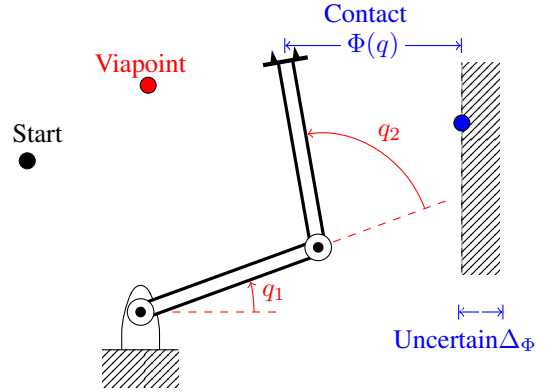


Fig. 1: Schematic: pass through a via-point and establish contact with a wall with uncertain location  $\Phi(q) + \Delta\Phi$ .

We analyze the problem of a 2-DOF manipulator as shown in Fig. 1. It allows us to show in a simple setting the main properties of the algorithm. The equations of motion are

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}) = \boldsymbol{\tau} + \mathbf{J}(\mathbf{q})^T \mathbf{F} \quad (32)$$

The vector  $\mathbf{q} = [q_1, q_2]^T$  contains the joints angular position.  $\mathbf{M}(\mathbf{q})$  is the inertia matrix,  $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})$  the vector of Coriolis and centrifugal forces,  $\mathbf{J}(\mathbf{q})$  is the end-effector Jacobian,  $\mathbf{F} \in \mathbb{R}^2$  the external forces and  $\boldsymbol{\tau} \in \mathbb{R}^2$  the input torques. The system dynamics can be easily written in the form

$$d\mathbf{x} = (\mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\boldsymbol{\tau})dt + \mathbf{G}(\mathbf{x})d\mathbf{w}$$

where we include additive process uncertainty  $d\mathbf{w}$  and denote the state as  $\mathbf{x} = [\mathbf{q}^T, \dot{\mathbf{q}}^T]^T$ , and the measurement model is

$$\mathbf{y} = [q_1 \quad q_2 \quad \dot{q}_1 \quad \dot{q}_2]^T + \boldsymbol{\gamma}$$

where  $\boldsymbol{\gamma}$  is Gaussian noise with variance  $\Gamma$ .

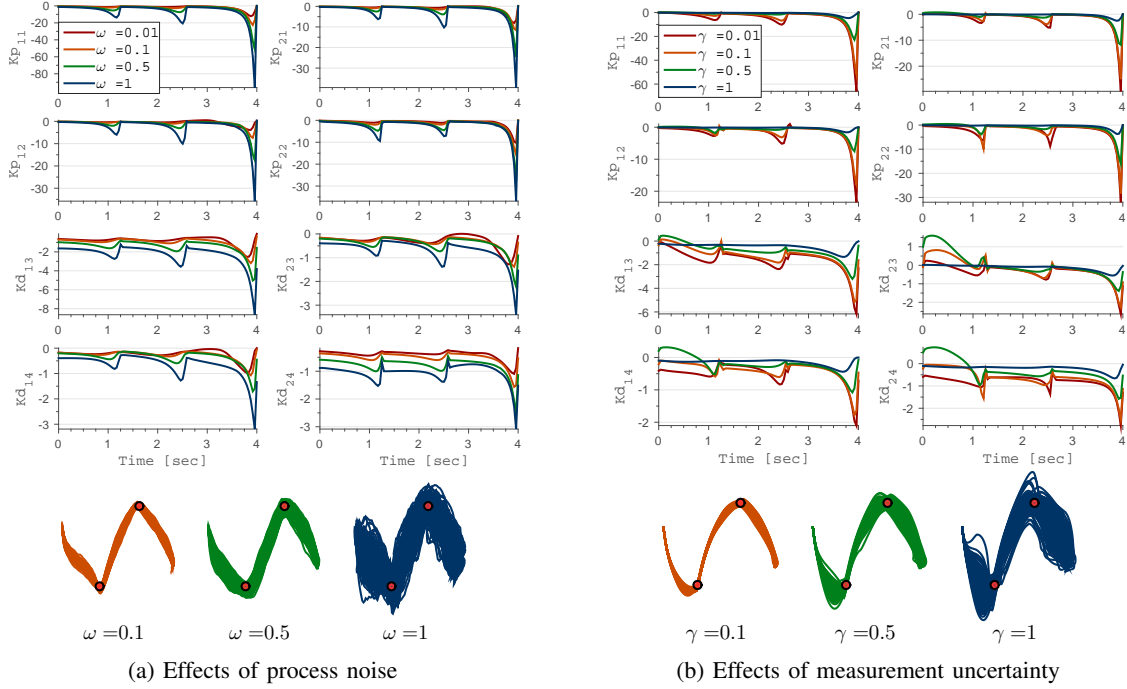


Fig. 2: Comparison between process noise and measurement uncertainty. The control gains for various level of noise are shown in the upper graphs. Sample trajectories are shown in the lower graphs for both varying process noise (left) and measurement uncertainty (right). The red dots represent the via points. In all the experiments the risk sensitive parameter  $\sigma = 2.5$ .

#### A. Experiment 1: Process Noise vs. Measurement Uncertainty

We compare the effect of process and measurement noise in the control law in a motion task between two points with two way-points. Task performance is measured by

$$\mathcal{J} = \sum_0^{t_f} c_u \tau^T \tau + \sum_{i=1}^{N_{via}} c_i \log(\cosh(\|x - x_i\|_2)) + c_{t_f} \log(\cosh(\|x - x_{t_f}\|_2))$$

where  $x, x_i, x_{t_f} \in \mathbb{R}^4$  are the current, viapoints and final desired end-effector positions respectively.  $c_u, c_i, c_{t_f}$  are cost weights. The nonlinear cost  $\log(\cosh(\cdot))$  is a soft absolute value to demonstrate that general nonlinear costs can be used.

We first evaluate the effects of increasing process noise under no measurement uncertainty. In Fig. 2a feedback gains for the motion task under several noise intensities are shown. In general, they are higher for regulating behavior at the viapoints and goal position. As process noise increases, the cost of uncertainty does too, because we might miss the viapoints or the goal due to disturbances. This can be seen in sample trajectories (Fig. 2a), where the variance of the trajectories due to noise has increased. In this case, the trade-off between cost of uncertainty and control-effort involves feedback gains proportional to the process noise, **the higher the process noise, the higher the feedback gains**.

In a second set of simulation, we test the effect of increasing measurement uncertainty under no process noise. In Fig. 2b, feedback gains and sample trajectories for different values of measurement noise are shown. Feedback gains are also higher close to viapoints and goal position, and sample trajectories are similar to the ones with process noise. The

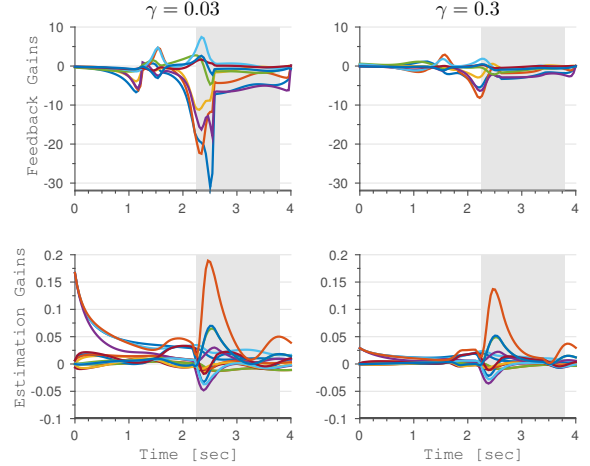


Fig. 3: Feedback and Estimation Gains for a motion-force task.

big difference is that the optimal control solution for this case is to trust feedback proportionally to the information content of the measurements, namely, **the higher the measurement uncertainty, the lower the feedback gains**. Fig. 2b shows how under low measurement noise, feedback control using high gains is possible and optimal. But under high measurement noise, a low impedance control law is better. We note that during the evaluation of the controller online estimation is used as it achieves better performance than using the precomputed sequence of estimation gains. In these experiments, we kept the risk sensitive parameter constant as it is not the focus of this paper (see for example [1]). However, the effects of process and measurement noise are qualitatively similar for all allowed values of  $\sigma$  (data not shown).



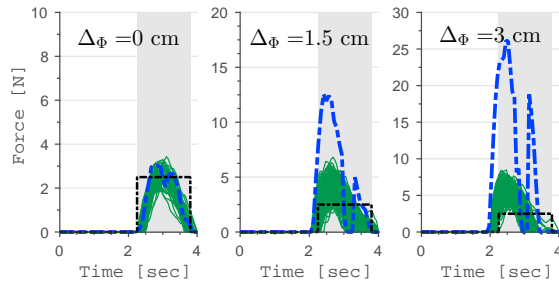


Fig. 4: Contact forces given a perturbation at the contact location. Black dashed lines show reference desired forces and the dashed blue lines show the forces using a controller optimized with process noise but low measurement noise.

### B. Experiment 2: Establishing a contact with the environment

In this experiment, the robot needs to pass through a via-point and then make contact with a wall at an uncertain location ( $\Delta_\Psi$ ), as shown in Fig. 1. While still simple to be carefully analyzed, the experiment addresses the role of measurement uncertainty when interacting with an uncertain environment, which is important for manipulation and locomotion tasks. Performance is measured by

$$\mathcal{J} = \sum_0^t c_u \tau^T \tau + c_{via} \log(\cosh(\|x_{t_{via}} - x_{via}\|_2)) + \sum_{t_{cnt_0}}^{t_{cnt_f}} c_{cnt} \log(\cosh(\Phi(q)) \cosh(\|\mathbf{F} - \mathbf{F}_{des}\|_2))$$

$\mathbf{F}_{des} \in \mathbb{R}^2$  is the desired force,  $\Phi(q)$  the distance to the contact,  $c_{cnt}$ ,  $c_{via}$ ,  $c_u$  are cost weights. This cost rewards low torques, passing a via-point  $x_{via}$  at time  $t_{via}$ , making contact  $\Phi(q) = 0$  and exerting a desired force from  $t_{cnt_0}$  to  $t_{cnt_f}$  (shown as shaded areas in Figs. 3-4). The external force  $\mathbf{F}$  is modeled as a stiff spring and is part of the dynamic model such that its effect is known to the optimizer.

Fig. 3 shows feedback and estimation gains for two measurement noise values  $\gamma$  that encodes uncertainty in the state, out of which distance to contact is computed. Feedback gains show two peaks around 1 and 2.2s, when passing the viapoint and when contact happens. Feedback gains for  $\gamma = 0.03$  are higher than for  $\gamma = 0.3$ , where control is more cautious. Estimation gains show a similar behavior, they are higher for low measurement noise. Interestingly, passing the viapoint does not affect them but when the contact is expected, they are higher because contact provides location information.

Fig. 4 shows force profiles of contact interaction with the wall ( $\gamma = 0.3$ ). Here we also show the force profiles with a control law using process noise ( $\omega = 0.2$ ) but low measurement noise ( $\gamma = 0.003$ ). We see that with the controller using measurement uncertainty when contact happens before it was expected ( $\Delta_\Phi = 1.5$  or  $\Delta_\Phi = 3.0$  cm), forces are higher, but the interaction is not as aggressive as it would be with higher feedback gains. In the case of process noise, since the feedback gains are higher we see much higher contact forces (blue lines) and even a loss of contact (right graph in Fig. 4).

These results illustrate a potentially useful behavior: policies sensitive to measurement uncertainty lead to low impedance behavior in face of too high uncertainty. In a receding horizon

setting, the impedance behavior would then be adapted as the robot gains more information about the state of the environment (e.g. after making a contact). The execution does not exploit sensed contact forces, which could improve further the dynamic interaction, however it is able to find a feedback control policy that can safely interact with the environment, despite uncertainty in the position of the wall.

## VI. DISCUSSION

In our experiment, we have been shown that process noise is fundamentally different from measurement noise. While the first one is a dynamics disturbance that requires high feedback gain control; the second one represents uncertainty in information about the state, and can use high feedback gains only under highly-informative measurements.

The fundamental difference between process and measurement noise effects on the control law comes from the cost they penalize. Cost of uncertainty due to process noise increases with terms of the form  $\mathbf{C}_t \Omega_t \mathbf{C}_t^T$ . If there is no control action, the process noise increases the cost. Therefore, regulation with high gains is optimal. For measurement noise, cost increases with terms  $\mathbf{K}_t \mathbf{D}_t \Gamma_t \mathbf{D}_t^T \mathbf{K}_t^T$  and estimation gains are inversely proportional to measurement noise. Therefore, not making use of rich-information measurements has a cost and requires high feedback gains. For poor-information measurements, we incur very low cost and control with lower gains is optimal. This behavior can be exploited in robotic tasks with dynamic interactions. For example when making a contact, behaving compliant under poor contact-information is robust. Once the contact is established and position certainty is higher, feedback gains would then be increased. In a receding horizon setup, gain in information about the current state of the world after contact would allow to online adapt the feedback policy.

From a computational point of view, the algorithm should scale to more complex systems. We can approximate the complexity of a call to the dynamics with its heaviest computation (factorization and back-substitution of  $M(q)$ ) as roughly  $O(n^3)$ ,  $n$  being the number of states. The most expensive computation is that of first derivatives  $O(Nn^4)$ ,  $N$  being the number of timesteps in the horizon. This is in the same order of complexity as other iterative approaches that show very good performance on more complicated robotic tasks [10, 11], but however did not exploit measurement uncertainty for control.

## VII. CONCLUSION

We have presented an iterative algorithm for finding locally-optimal feedback controllers for nonlinear systems with additive measurement uncertainty. In particular we showed that measurement uncertainty leads to very different behaviors than process noise and it can be exploited to create low impedance behaviors in uncertain environments (e.g. during contact interaction). This opens the possibility for planning and controlling contact interactions robustly based on controllers sensitive to measurement noise. In a receding horizon setting, it could be possible to regulate impedance in a meaningful way depending on the current uncertainty about the environment.



## ACKNOWLEDGMENTS

This research was supported by the Max-Planck Society, the European Research Council under the European Unions Horizon 2020 research and innovation programme (grant agreement No 637935), and the Max Planck ETH Center for Learning Systems.

## REFERENCES

- [1] Farbod Farshidian and Jonas Buchli. Risk Sensitive, Nonlinear Optimal Control: Iterative Linear Exponential-Quadratic Optimal Control with Gaussian Noise. 2015. URL <http://arxiv.org/abs/1512.07173>.
- [2] D. Jacobson. Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games. *IEEE Transaction on Automatic Control*, 18(2):124–131, 1973. doi: 10.1109/TAC.1973.1100265. URL <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1100265>.
- [3] MR. James, JS. Baras, and LJ. Elliot. Risk-sensitive control and dynamic games for partially observed discrete-time nonlinear systems. *IEEE Transactions on Automatic Control*, 39(4):780–792, 1994.
- [4] Weiwei Li and Emmanuel Todorov. Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic system. *International Journal of Control*, 80(9):1439 – 1453, 2007.
- [5] David Mayne. A Second-order Gradient Method for Determining Optimal Trajectories of Non-linear Discrete-time Systems. *International Journal of Control*, 3(1): 85–95, 1966. doi: 10.1080/00207176608921369.
- [6] Stefan Schaal and Atkeson C. Learning control in robotics. *Robotics and Automation Magazine*, 17:20–29, 2010. doi: 10.1109/MRA.2010.936957.
- [7] Athanasios Sideris and James Bobrow. An efficient sequential linear quadratic algorithm for solving nonlinear optimal control problems. *IEEE Trans. Automat. Contr.*, 50(12):2043–2047, 2005.
- [8] Jason Speyer, John Deyst, and D. Jacobson. Optimization of stochastic linear systems with additive measurement and process noise using exponential performance criteria. *IEEE Transaction on Automatic Control*, 19(4):358–366, 1974. doi: 10.1109/TAC.1974.1100606. URL <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1100606>.
- [9] Charlotte Striebel. Sufficient statistics in the optimum control of stochastic systems. *Journal of Mathematical Analysis and Applications*, 12(3):576 – 592, 1965. doi: [http://dx.doi.org/10.1016/0022-247X\(65\)90027-2](http://dx.doi.org/10.1016/0022-247X(65)90027-2).
- [10] Yuval Tassa, Tom Erez, and E Todorov. Fast model predictive control for reactive robotic swimming.
- [11] Yuval Tassa, Nicolas Mansard, and Emo Todorov. Control-limited differential dynamic programming. In *2014 IEEE International Conference on Robotics and Automation, ICRA 2014, Hong Kong, China, May 31 - June 7, 2014*, pages 1168–1175, 2014. doi: 10.1109/ICRA.2014.6907001. URL <http://dx.doi.org/10.1109/ICRA.2014.6907001>.
- [12] Emmanuel Todorov. Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural Computation*, 17(5): 1084–1108, 2005. doi: 10.1162/0899766053491887.
- [13] Emmanuel Todorov and Weiwei Li. A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems. *American Control Conference, 2005. Proceedings of the 2005*, 1: 300 – 306, 2005. doi: 10.1109/acc.2005.1469949.
- [14] P. Whittle and J. Kuhn. A hamiltonian formulation of risk-sensitive linear quadratic gaussian control. *International Journal on Control*, 43:1–12, 1986.